# Soft-Sensing for the Feedwater Flowrate at PWRs Using a GMDH Algorithm

Dong Su Kim [a], Sung Han Lee [a], Man Gyun Na [a*] and Dong Hyuk Lim [b]
[a] *Department of Nuclear Engineering, Chosun University, 375 Seosuk-dong, Dong-gu, Gwangju 501-759, Korea*
[b] *Korea Atomic Energy Research Institute, Dukjin-Dong, Yuseong-Gu, Daejeon*
**Corresponding author: magyna@chosun.ac.kr*

## 1. Introduction

Currently, Venturi flow meters are being used to measure the feedwater flowrate in most pressurized water reactors (PWRs). These meters can induce measurement drift owing to corrosion product buildup near the meter orifice through long-term operation. Fouling of the Venturi meter decreases the accuracy of the existing hardware sensors. These requirements result in nuclear power plants operating at lower-than-planned power levels. Therefore, considerable research has been focused on resolving the inaccurate measurement issue of the feedwater flowrate. This study employed the Group Method of Data Handling (GMDH) which is one of the data driven models, such as Artificial Neural Networks (ANNs), to increase the thermal efficiency by accurately estimating online the feedwater flowrate. The GMDH method was trained using the informative data selected through a subtractive clustering (SC) scheme. In addition, the uncertainty of the GMDH algorithm was analyzed using a statistical uncertainty method.

## 2. Soft-Sensing for Feedwater Flowrate

### *2.1 Group Method of Data Handling*

A GMDH model was developed for soft-sensing of the feedwater flowrate of a PWR. The GMDH algorithm can generally find interrelations in data automatically to improve the prediction accuracy and select the optimal structure of the model or networks.

The GMDH algorithm employs a self-organizing modeling algorithm with the flexibility of selecting nonlinear forms of the basic inputs.

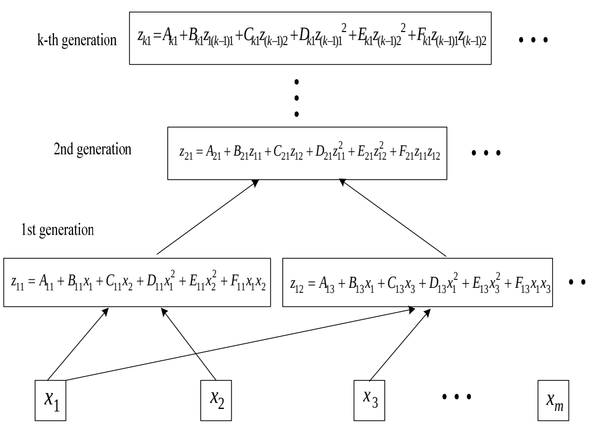


Fig. 1. Branch structure of the GMDH model [1]

Fig. 1 shows the branch structure of the GMDH algorithm [1], which begins with the basic inputs at the first (bottom) level, and becomes more complex as the number of layers increases.

The original GMDH method employed the following general form at each level of the successive approximation [1]:

$$y = f(x_i, x_j) = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_i x_j \qquad (1)$$

The GMDH algorithm employs a high-order polynomial in the Kolmogorov-Gabor form [2]. The Kolmogorov-Gabor (called as Ivakhnenko polynomial) is expressed as follows:

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k \ldots \qquad (2)$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_m)$ is an input variable vector and $\mathbf{a} = (a_0, a_i, a_{ij}, a_{ijk,} \cdots)$ is a vector of coefficients or a weight of the Kolmogorov-Gabor polynomial. Components of the input vector $\mathbf{x}$ can be independent variables, functional forms or finite difference terms.

The algorithm generates and tests all input-output combinations in a system. Each element of the system executes a function of two inputs. The variables of the elements are normally decided using a regression technique. A type of threshold set at each generation determines if the outputs of the elements in a generation are acceptable. The output of an element is eliminated in a current generation when the result is larger than the threshold value. Those variables or elements that are useful in predicting the proper output are filtered out. The generations are repeated until satisfactory results are obtained.

### *2.2 Training Data Selection*

A GMDH model can be well trained using informative data. It is expected that the input and output training data exist in the form of clusters and the data at these cluster centers is more informative than the neighboring data. In this paper, the cluster centers were determined using a subtractive clustering (SC) scheme and were used as the training data set.

In general, after the $i$ th cluster center has been determined, the potential of each data point is revised using the following equation:

$$P_{i+1}(k) = P_i(k) - P^*(i)e^{-4\left\|\mathbf{x}_k - \mathbf{x}_i^*\right\|^2 / r_\beta^2}, \ k = 1, 2, \ldots, n \qquad (3)$$

where $\mathbf{x}_i^*$ is the location of the $i-th$ cluster center and $P^*(i)$ is its potential value.

These calculations stop if the inequality $P^*(i) < \varepsilon P^*(1)$ is true, else the calculations are repeated. The input/output data positioned in the cluster centers was selected in order to train the GMDH model.

### 3. Uncertainty Analysis

A statistical uncertainty analysis method was used in the present study. Statistical uncertainty analysis generates many bootstrap samples of training and checking data sets initially and retrains the GMDH model parameters on each bootstrap sample. After repetitive sampling and training, the values predicted by the GMDH models provide a distribution. This distribution is then used to calculate the prediction intervals. The bootstrap pairs sampling algorithm was used in this study. The available data was divided into development data and test data. After the test data was selected initially, the data remaining becomes the development data consisting of a large pool of data from which training and checking samples are drawn.

Generate $J$ (training and checking) samples ($J = 100$, the number of bootstrap samples, in this paper) from the development data, each one of size $n$ is drawn with a replacement. The $j-th$ sample is denoted by

$$\left\{(\mathbf{x}_1^j, y_1^j), (\mathbf{x}_2^j, y_2^j), \cdots, (\mathbf{x}_l^j, y_l^j), \cdots, (\mathbf{x}_n^j, y_n^j)\right\} \quad (4)$$

Estimate the variance and bias of a prediction $\hat{y}_0$ at an observation data point $\mathbf{x}_0$ by

$$Var(\hat{y}_0) = \frac{1}{J-1}\sum_{j=1}^{J}\left[\hat{y}_0^j - \bar{\hat{y}}_0\right]^2 \text{ where } \bar{\hat{y}}_0 = \frac{1}{J}\sum_{j=1}^{J}\hat{y}_0^j \quad (5)$$

$$bias = \left\{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{J}\sum_{j=1}^{J}\left[\hat{y}_k^j - y_k^j\right]^2\right\}^{1/2} \quad (6)$$

where $K$ is the number of development data points.

The estimate with a 95% confidence interval for an arbitrary test input $\mathbf{x}_o$ is determined by the following:

$$\hat{y}_0 \pm 2\sqrt{Var(\hat{y}_0) + bias^2} = \hat{y}_0 \pm \delta \quad (7)$$

### 4. Application of the Soft-Sensing Model to the Feedwater Flow Prediction

The proposed algorithm was confirmed using the real plant startup data of the Yonggwang Nuclear Power Plant Unit 3 (YGN3). The data consisted of sixteen signals measured from the primary and secondary systems of the nuclear power plant, focusing on the steam generator (SG). The acquired SG feedwater flowrate was the target output signal of the GMDH model and all other signals were available inputs for the GMDH model.

In this paper, the acquired real plant data was divided into three types of data to train, check and test the GMDH model; training data, checking data and test data.
Table 1 summarizes the performance results of the GMDH model and compares the performances of the GMDH model with other two data-based models [2], such as fuzzy inference system (FIS) and support vector regression (SVR). The GMDH model does not outperform the FIS and SVR models for the training and verification data but outperforms the other models for the test data of which the result is more important.

Table 1: Performance of the data-based models

| Model | Data type | RMS error (%) | Relative Maximum error(%) | Number of data points |
|---|---|---|---|---|
| GMDH | Training data | 0.1832 | 2.6978 | 1000 |
| | Verification data | 0.1408 | 2.6978 | 1800 |
| | Test data | 0.0581 | 0.4098 | 201 |
| FIS [2] | Training data | 0.1018 | 1.2092 | 1000 |
| | Verification data | 0.0838 | 1.2092 | 1800 |
| | Test data | 0.3181 | 3.7025 | 201 |
| SVR [2] | Training data | 0.2105 | 1.4278 | 1000 |
| | Verification data | 0.1896 | 1.4278 | 1800 |
| | Test data | 0.2085 | 1.2497 | 201 |

### 5. Conclusions

In this study, the GMDH model was used as a soft-sensing algorithm to accurately predict the feedwater flowrate. The proposed GMDH model was applied and verified using the acquired real plant startup data of YGN3. In addition, more informative data obtained from an SC scheme was applied to train the GMDH model. The GMDH model was compared with the SVR and FIS models developed in a previous study [2]. The relative RMS error and maximum error of the GMDH model for the test data was 0.0581% and 0.4098%, respectively. The performance of the GMDH model was superior or similar to those of SVR and FIS models. In addition, through statistical uncertainty analysis, estimates with a 95% confidence interval were obtained for the 201 test data points. Since the prediction interval is small, it is expected that the GMDH model can be applied successfully to validate and monitor existing feedwater flow meters.

### REFERENCES

[1] S.J. Farlow, "Self-Organizing Methods in Modeling: GMDH Type Algorithms," Marcel Dekker, New York, 1984.
[2] H.Y. Yang, S.H. Lee, and M.G. Na, "Monitoring and Uncertainty Analysis of Feedwater Flow Rate Using Data-based Modeling Methods," *IEEE Trans. Nucl. Sci.*, Vol. 56, No. 4, pp. 2426-2433, Aug. 2009.